The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Artificial Intelligence Applied to Assistive Technologies and Smart Environments:
Technical Report WS-16-01

Deep Activity Recognition Models with Triaxial Accelerometers

**Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato,
Linda Doyle, Shaowei Lin, Hwee-Pink Tan**

Abstract

Despite the widespread installation of accelerometers in almost all mobile phones and wearable devices, activity recognition using accelerometers is still immature due to the poor recognition accuracy of existing recognition methods and the scarcity of labeled training data. We consider the problem of human activity recognition using triaxial accelerometers and deep learning paradigms. This paper shows that deep activity recognition models (a) provide better recognition accuracy of human activities, (b) avoid the expensive design of handcrafted features in existing systems, and (c) utilize the massive unlabeled acceleration samples for unsupervised feature extraction. We show substantial recognition improvement on real world datasets over state-of-the-art methods of human activity recognition using triaxial accelerometers.

Introduction

Existing sensor-based activity recognition systems (Chen et al. 2012) use shallow and conventional supervised machine learning algorithms such as multilayer perceptrons (MLPs), support vector machines, and decision trees. This reveals a gap between the recent developments of deep learning algorithms and existing sensor-based activity recognition systems. When deep learning is applied for sensor-based activity recognition, it results in many advantages in terms of system performance and flexibility. Firstly, deep learning provides an effective tool for extracting high-level feature hierarchies from high-dimensional data which is useful in classification and regression tasks (Salakhutdinov 2015). These automatically generated features eliminate the need for handcrafted features of existing activity recognition systems. Secondly, deep generative models, such as deep belief networks (Hinton, Osindero, and Teh 2006), can utilize unlabeled activity samples for model fitting in an unsupervised pre-training phase which is exceptionally important due to the scarcity of labeled activity datasets. In the contrary, unlabeled activity datasets are abundant and cheap to collect. Thirdly, deep generative models are more robust against the overfitting problem as compared to discriminative models (e.g., MLP) (Mohamed, Dahl, and Hinton 2012).

In this paper, we present a systematic approach towards detecting human activities using deep learning and triaxial accelerometers. This paper is also motivated by the success of deep learning in acoustic modeling (Mohamed, Dahl, and Hinton 2012; Dahl et al. 2012), as we believe that speech and acceleration data have similar patterns of temporal fluctuations. Our approach is grounded over the automated ability of deep activity recognition models in extracting intrinsic features from acceleration data. Our extensive experiments are based on three public and community-based datasets. In summary, our main results on deep activity recognition models can be summarized as follows:

- **Deep versus shallow models**. Our experimentation shows that using deep activity recognition models significantly enhances the recognition accuracy compared with conventional shallow models. Equally important, deep activity recognition models automatically learn meaningful features and eliminate the need for the hand-engineering of features, e.g., statistical features, in state-of-the-art methods.
- **Exploiting unlabeled data**. The scarce availability of labeled activity data motivates the exploration of semi-supervised learning techniques for a better fitting of activity classifiers. Our experiments show the importance of the generative (unsupervised) training of deep activity recognition models in weight tuning and optimization.
- **Spectrogram analysis**. Accelerometers generate multi-frequency, aperiodic, and fluctuating signals which complicate the activity recognition using time series data. We show that using spectrogram signals instead of the raw acceleration data exceptionally helps the deep activity recognition models to capture variations in the input data.

Related Work

In this section, we will focus on classification and feature engineering methods for activity recognition using accelerometers. For a more comprehensive review of the field, we refer interested readers to recent survey papers (Lara and Labrador 2013; Chen et al. 2012).

Limitations of Shallow Classifiers

Machine learning algorithms have been used for a wide range of activity recognition applications (Parkka et al. 2006; Khan et al. 2010; Altun and Barshan 2010; Kwapisz, Weiss, and Moore 2011), allowing the mapping between feature sets and various human activities. The classification of accelerometer samples into static and dynamic activities using MLPs is presented in (Khan et al. 2010). Conventional neural networks, including MLPs, often stuck in local optima (Rumelhart, Hinton, and Williams 1986) which leads to poor performance of activity recognition systems. Moreover,

training MLPs using backpropagation (Rumelhart, Hinton, and Williams 1986) only hinders the addition of many hidden layers due to the vanishing gradient problem. The authors in (Parkka et al. 2006) used decision trees and MLPs to classify daily human activities. In (Berchtold et al. 2010), a fuzzy inference system is designed to detect human activities. (Kwapisz, Weiss, and Moore 2011) compared the recognition accuracy of decision tree (C4.5), logistic regression, and MLPs, where MLPs are found to outperform the other methods.

In this paper, we show significant recognition accuracy improvement on real world datasets over state-of-the-art methods for human activity recognition using triaxial accelerometers. Additionally, even though some previous works have purportedly reported promising results of activity recognition accuracy, they still require a degree of handcrafted features as discussed below.

Limitations of Handcrafted Features

Handcrafted features are widely utilized in existing activity recognition systems for generating distinctive features that are fed to classifiers. The authors in (Altun and Barshan 2010; Berchtold et al. 2010; Kwapisz, Weiss, and Moore 2011; Xu et al. 2012; Catal et al. 2015) utilized statistical features, e.g., mean, variance, kurtosis and entropy, as distinctive representation features. On the negative side, statistical features are problem-specific, and they poorly generalize to other problem domains. In (Zappi et al. 2008), the signs of raw signal (positive, negative, or null) are used as distinctive features. Despite its simple design, these sign features are plain and cannot represent complex underlying activities which increase the number of required accelerometer nodes. The authors in (Bächlin et al. 2010) used the energy and frequency bands in detecting the freezing events of Parkinson's disease patients. Generally speaking, any handcrafted-based approach involves laborious human intervention for selecting the most effective features and decision thresholds from sensory data.

Quite the contrary, data-driven approaches, e.g., using deep learning, can learn discriminative features from historical data which is both systematic and automatic. Therefore, deep learning can play a key role in developing self-configurable framework for human activity recognition. The author in (Plötz, Hammerla, and Olivier 2011) discussed the utilization of a few feature learning methods, including deep learning, in activity recognition systems. Nonetheless, this prior work is elementary in its use of deep learning methods, and it does not provide any analysis of the deep network construction, e.g., setup of layers and neurons.

Problem Statement

This section gives a formal description of the activity recognition problem using accelerometer sensors.

Data Acquisition

Consider an accelerometer sensor that is attached to a human body and takes samples (at time index t) of the form

$$\mathbf{r}_t = \mathbf{r}_t^* + \mathbf{w}_t, \quad t = 1, 2, \dots \tag{1}$$

where $\mathbf{r}_t = \begin{bmatrix} r_t^x & r_t^y & r_t^z \end{bmatrix}^T$ is a 3D accelerometer data point generated at time t and composed of r_t^x, r_t^y, and r_t^z which are the x-acceleration, y-acceleration, and z-acceleration components, respectively. The proper acceleration in each axis channel is a floating-point value that is bounded to some known constant $B > 0$ such that $|r_t^x| \leq B$, $|r_t^y| \leq B$, and $|r_t^z| \leq B$. For example, an accelerometer with $B = 2$g units indicates that it can record proper acceleration up to twice the gravitational acceleration (recall that $1\text{g} \simeq 9.8 \frac{meter}{second^2}$). Clearly, an accelerometer that is placed on a flat surface record a vertical acceleration value of ± 1g upward. $\mathbf{r}_t^* \in \mathbb{R}^3$ is a vector that contains 3-axial noiseless acceleration readings. $\mathbf{w}_t \in \mathbb{R}^3$ is a noise vector of independent, zero-mean Gaussian random variables with variance σ_w^2 such that $\mathbf{w}_t \backsim \mathcal{N}(0, \sigma_w^2 \mathbb{I}_3)$. Examples of added noise during signal acquisition include the effect of temperature drifts and electromagnetic fields on electrical accelerometers (Fender et al. 2008).

Three channel frames $\mathbf{s}_t^x$, $\mathbf{s}_t^y$, and $\mathbf{s}_t^z \in \mathbb{R}^{\mathrm{N}}$ are then formed to contain the x-acceleration, y-acceleration, and z-acceleration components, respectively. Particularly, these channel frames are created using a sliding window as follows:

$$\mathbf{s}_t^x = \begin{bmatrix} r_t^x & \cdots & r_{t+N-1}^x \end{bmatrix}^T, \tag{2}$$

$$\mathbf{s}_t^x = \begin{bmatrix} r_t^y & \cdots & r_{t+N-1}^y \end{bmatrix}^T, \tag{3}$$

$$\mathbf{s}_t^z = \begin{bmatrix} r_t^z & \cdots & r_{t+N-1}^z \end{bmatrix}^T. \tag{4}$$

The sequence size N should be carefully selected such as to ensure an adequate and efficient activity recognition. We assume that the system supports M different activities. Specifically, let $\mathcal{A} = \{a_1, a_2, \dots, a_M\}$ be a finite activity space. Based the windowed excerpts $\mathbf{s}_t^x$, $\mathbf{s}_t^x$, and $\mathbf{s}_t^z$, the activity recognition method infers the occurrence of an activity $y_t \in \mathcal{A}$.

Data Preprocessing

A *spectrogram* of an accelerometer signal is a three dimensional representation of changes in the acceleration energy content of a signal as a function of frequency and time. Historically, spectrograms of speech waveforms are widely used as distinguishable features in acoustic modeling, e.g., the mel-frequency cepstral (Zheng, Zhang, and Song 2001). In this paper, we use the spectrogram representation as the input of deep activity recognition models as it introduces the following advantages:

1. **Classification accuracy**. The spectrogram representation provides interpretable features in capturing the intensity differences among nearest acceleration data points. This enables the classification of activities based on the variations of spectral density which reduce the classification complexity.
2. **Computational complexity**. After applying the spectrogram on $\mathbf{s}_t^x$, $\mathbf{s}_t^x$, and $\mathbf{s}_t^z$, the length of the spectral signal is $L = 3(\frac{N}{2} + 1)$ while the time domain signal length is $3N$. This significantly reduces the computational burdens of any classification method due to the lower data dimensionality.

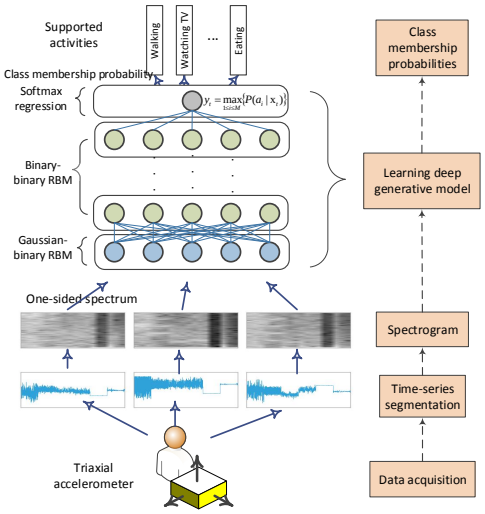

Figure 1: Activity recognition using deep activity recognition model. Our system automatically (1) takes triaxial acceleration time series, (2) extracts the spectrogram of windowed excerpts, (3) computes intrinsic features using a deep generative model, and then (4) recognizes the underlying human activities by finding the posterior probability distribution $\{P(a_i|\mathbf{x}_t)\}_{i=1}^{M}$. This deep architecture outperforms existing methods for human activity recognition using accelerometers as shown by the experimental analysis on real world datasets.

Henceforth, the spectrogram signal of the triaxial accelerometer is denoted as $\mathbf{x}_t \in \mathbb{R}^L$, where $L = 3(\frac{N}{2}+1)$ is the concatenated spectrogram signals from the triaxial input data.

Deep Learning for Activity Recognition: System and Model

Our deep model learns not only the classifier's weights used to recognize different activities, but also the informative features for recognizing these activities from raw data. This provides a competitive advantage over traditional systems that are hand-engineered. The model fitting and training consist of two main stages: (i) An unsupervised, generative, and pre-training step, and (ii) a supervised, discriminative, and fine-tuning step. The pre-training step generates intrinsic features based on a layer-by-layer training approach using unlabeled acceleration samples only. Firstly, we use deep belief networks (Hinton, Osindero, and Teh 2006) to find the activity membership probabilities.

Figure 1 shows the working flow of the proposed activity recognition system. We implement deep activity recognition models based on deep belief networks (BBNs). DBNs are generative models composed of multiple layers of hidden units. In (Hinton, Osindero, and Teh 2006), the hidden units are formed from restricted Boltzmann machines (RBMs) which are trained in a layer-by-layer fashion. Notably, an alternative approach is based on using stacked autoencoders (Bengio et al. 2007). An RBM is a bipartite graph that is restricted in that no weight connections exist between hidden units. This restriction facilitates the model fitting as the hidden units become conditional independent for a given input vector. After the unsupervised pre-training, the learned weights are fine-tuned in an up-down manner using available data labels. A practical tutorial on the training of RBMs is presented in (Hinton 2012).

Deep Activity Recognition Models

DBNs (Hinton, Osindero, and Teh 2006) can be trained on greedy layer-wise training of RBMs as shown in Figure 2. In our model, the acceleration spectrogram signals $\mathbf{x}$ are continuous and are fed to a deep activity recognition model. As a result, the first layer of the deep model is selected as a Gaussian-binary RBM (GRBM) which can model the energy content in the continuous accelerometer data. Afterward, the subsequent layers are binary-binary RBMs (BRBMs). RBMs are energy-based probabilistic models which are trained using stochastic gradient descent on the negative log-likelihood of the training data. For the GRBM layer, the energy of an observed vector $\mathbf{v} = \mathbf{x}$ and a hidden code $\mathbf{h}$ is denoted as follows:

$$E(\mathbf{v}=\mathbf{x},\mathbf{h}) = \frac{1}{2}(\mathbf{v}-\mathbf{b})^{\top}(\mathbf{v}-\mathbf{b}) - \mathbf{c}^{\top}\mathbf{h} - \mathbf{v}^{\top}\mathbf{W}\mathbf{h} \quad (5)$$

where $\mathbf{W}$ is the weight matrix connecting the input and hidden layers, $\mathbf{b}$ and $\mathbf{c}$ are the visible and hidden unit biases, respectively. For a BRBM, the energy function is defined as follows:

$$E(\mathbf{v},\mathbf{h}) = -\mathbf{b}^{\top}\mathbf{v} - \mathbf{c}^{\top}\mathbf{h} - \mathbf{v}^{\top}\mathbf{W}\mathbf{h}. \quad (6)$$

An RBM can be trained using the contrastive divergence approximation (Hinton 2002) as follows:

$$\triangle W_{ij} = \alpha\left(\langle v_i h_j\rangle_{\text{data}} - \langle v_i h_j\rangle_1\right) \quad (7)$$

where α is a learning rate. $\langle v_i h_j\rangle_{\text{data}}$ is the expectation of reconstruction over the data, and $\langle v_i h_j\rangle_1$ is the expectation of reconstruction over the model using one step of the Gibbs sampler. Please refer to (Hinton, Osindero, and Teh 2006; Hinton 2012) for further details on the training of DBNs. For simplicity, we denote the weights and biases of a DBN model as θ which can be used to find the posterior probabilities $P(a_i|\mathbf{x}_t,\theta)$ for each joint configuration $(a_i,\mathbf{x}_t)$.

To this end, the underlying activity y_t can be predicted at time t using the softmax regression as follows:

$$y_t = \arg\max_{1\le i\le M}\{P(a_i|\mathbf{x}_t,\theta)\}. \quad (8)$$

Computational Complexity

Our algorithm consists of three working phases: (a) data gathering, (b) offline learning, and (c) online activity recognition and inference. The computational burden of the offline learning is relatively heavy to be run on a mobile device as it based on stochastic gradient descent optimization. Therefore, it is recommended to run the offline training of a deep activity recognition model on a capable server. Nonetheless, after the offline training is completed, the

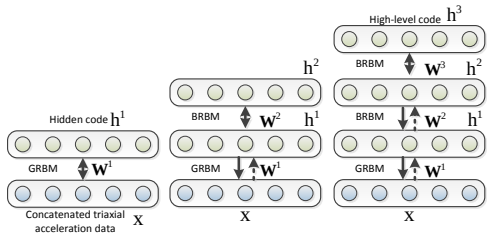

Figure 2: The greedy layer-wise training of DBNs. The first level is trained on triaxial acceleration data. Then, more RBMs are repeatedly stacked to form a deep activity recognition model until forming a high-level representation.

model parameter θ is only disseminated to the wearable device where the online activity recognition is lightweight with a linear time complexity ($\mathcal{O}(L \times D)$), where D is the number of layers in the deep activity recognition model. Here, the time complexity of the online activity recognition system represents the time needed to recognize the activity as a function of the accelerometer input length. The time complexity of finding the short-time Fourier transform (STFT) is $\mathcal{O}(L \log(L))$.

Baselines and Result Summary

Datasets

For empirical comparison with existing approaches, we use three public datasets that represent different application domains to verify the efficiency of our proposed solution. These three testbeds are described as follows:

- **WISDM Actitracker dataset** (Kwapisz, Weiss, and Moore 2011): This dataset contains $1,098,213$ samples of one triaxial accelerometer that is programmed to sample at a rate of 20 Hz. The data samples belong to 29 users and 6 distinctive human activities of walking, jogging, sitting, standing, and climbing stairs. The acceleration samples are collected using mobile phones with Android operating system.
- **Daphnet freezing of gait dataset** (Bächlin et al. 2010): We used this dataset to demonstrate the healthcare applications of deep activity recognition models. The data samples are collected from patients with the Parkinson's disease. Three triaxial accelerometers are fixed at patient's ankle, upper leg, and trunk with a sampling frequency of 64 Hz. The objective is to detect freezing events of patients. The dataset contains $1,140,835$ experimentation samples from 10 users. The samples are labeled with either "freezing" or "no freezing" classes.
- **Skoda checkpoint dataset** (Zappi et al. 2008): The 10 distinctive activities of this dataset belong to a car maintenance scenario in typical quality control checkpoints. The sampling rate is 98 Hz. Even though the dataset contains 20 nodes of triaxial accelerometers, it would be inconvenient and costly to fix 20 nodes to employee hands which can hinder the maintenance work. Therefore, we use one accelerometer node (ID # 16) for the experimental validation of deep models.

Performance Measures

For binary classification (experimentation on the Daphnet dataset), we use three performance metrics: Sensitivity (TPR) $= \frac{\text{TP}}{\text{TP+FN}}$, specificity (TNR) $= \frac{\text{TN}}{\text{TN+FP}}$, and accuracy (ACC) $= \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$ where TP, TN, FP, and FN mean true positive, true negative, false positive, and false negative, respectively. For multiclass classification of non-overlapping activities, which are based on the experimentation of the WISDM Actitracker and Skoda checkpoint datasets, the average recognition accuracy (ACC) is found as ACC $= \frac{1}{M}\sum_{i=1}^{M}\frac{\text{TP}_i+\text{TN}_i}{\text{TP}_i+\text{TN}_i+\text{FP}_i+\text{FN}_i}$, where M is the number of supported activities.

Baselines

Table 1 summarizes the main performance results of our proposed method and some previous solutions on using the three datasets. Deep activity recondition models introduce significant accuracy improvement over conventional methods. For example, it improves accuracy by 6.53% over MLPs and 3.93% over ensemble learning on the WISDM Actitracker dataset. Similarly, significant improvements are also reported for the Daphnet freezing of gait and Skoda checkpoint datasets. This summarized result shows that the deep models are both (a) effective in improving recognition accuracy over state-of-the-art methods, and (b) practical for avoiding the hand-engineering of features.

Experiments on Real Datasets

Spectrogram Analysis

Figure 3 shows triaxial time series and spectrogram signals of 6 activities of the WISDM Actitracker dataset. Clearly, the high frequency signals (a.k.a. AC components) belong to activities with active body motion, e.g., jogging and walking. On the other hand, the low frequency signals (a.k.a. DC components) are collected during semi-static body motions, e.g., sitting and standing. Thereby, these low frequency activities are only distinguishable by the accelerometer measurement of the gravitational acceleration.

Performance Analysis

In our experiments, the data is firstly centered to the mean and scaled to a unit variance. The deep activity recognition models are trained using stochastic gradient decent with mini-batch size of 75. For the first GBRM layer, the pre-training learning rate is set to 0.001 with pre-training epochs of 150. For next BRBM layers, the number of pre-training epochs is fixed to 75 with pre-training learning rate of 0.01. The fine-tuning learning rate is 0.1 and the number of fine-tuning epochs is 1000. For interested technical readers, Hinton (Hinton 2012) provides a tutorial on training RBMs with many practical advices on parameter setting and tuning.

Deep Model Structure Figure 4 shows the recognition accuracy on different DBN structures (joint configurations of number of layers and number of neurons per layer). Two important results are summarized as follows:

Table 1: Comparison of our proposed solution against existing methods in terms of recognition accuracy. C4.5 is a decision tree generation method.

DATASET	REFERENCE	SOLUTION	WINDOW SIZE	ACCURACY (%)
WISDM	(Kwapisz, Weiss, and Moore 2011)	C4.5	10 sec	85.1
	(Kwapisz, Weiss, and Moore 2011)	Logistic regression		78.1
	(Kwapisz, Weiss, and Moore 2011)	MLPs		91.7
	(Catal et al. 2015)	Ensemble learning		94.3
	Our solution	**Deep learning models**		**98.23**
Daphnet	(Bächlin et al. 2010)	Energy threshold on power spectral density (0.5sec)	4 sec	TPR: 73.1 and TNR: 81.6
	(Hammerla et al. 2013)	C4.5 and k-NNs with feature extraction methods	-	TPR and TNR $\sim$ 82
	Our solution	**Deep learning models**	4 sec	**TPR and TNR $\sim$ 91.5**
Skoda	(Zappi et al. 2008)	Hidden Markov models	-	Node 16 (86), nodes 20, 22 and 25 (84)
	Our solution	**Deep learning models**	4 sec	Node 16 (**89.38**)

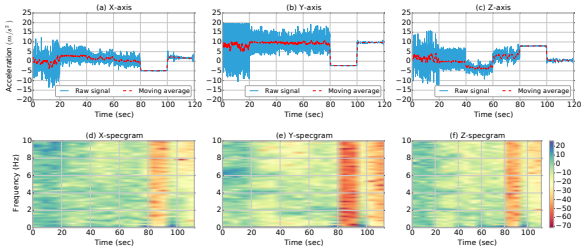

Figure 3: Frequency spectrum as a parametric representation. Data samples of a triaxial accelerometer and their corresponding spectrogram representation (WISDM Actitracker dataset). These samples belong to five everyday human activities: jogging $t \in [0, 20)$, walking $t \in [20, 40)$, upstairs $t \in [40, 60)$, downstairs $t \in [60, 80)$, sitting $t \in [80, 100)$, and standing $t \in [100, 120)$. The acceleration signal is usual subtle and only cover a small range of the frequency domain.

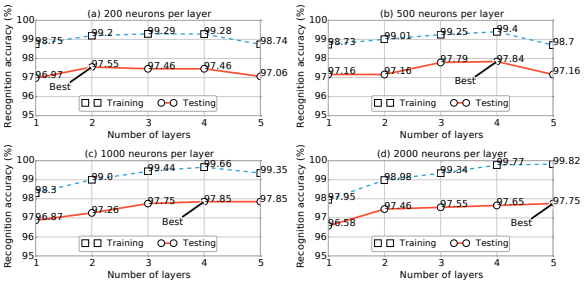

Figure 4: Optimizing deep activity recognition models. Activity recognition using the WISDM Actitracker dataset under different DBN setup. At each figure, the rates of activity recognition accuracy are shown for both training and testing data samples. The input length is 303 which corresponds to 10-second frames.

1. **Deep models outperforms shallow ones**. Clearly, the general trend in the recognition accuracy is that using more layers will enhance the recognition accuracy. For example, using 4 layers of 500 neurons at each layer is better than 2 layers of 1000 neurons at each layer, which is better than 1 layer of 2000 neurons.
2. **Overcomplete representations are advantageous**. An overcompete representation is achieved when the number of neurons at each layer is larger than the input length. An overcompete representation is essential for learning deep models with many hidden layers, e.g., deep model of 2000 neurons per layer. On the other hand, it is noted that a deep model will be hard to optimized when using undercomplete representations, e.g., 5 layers of 200 neurons at each layer. This harder optimization issue is distinguishable from the overfitting problem as the training data accuracy is also degrading by adding more layers (i.e., an overfitted model is diagnosed when the recognition accuracy on training data is enhancing by adding more layer while getting poorer accuracy on testing data). Therefore, we recommend 4x overcomplete deep activity recognition models (i.e., the number of neurons at each layer is four times the input size).

Pre-training Effects Table 2 shows the recognition accuracy of the WISDM Actitracker dataset with and without the pre-training phase. These results confirm the importance of the generative pre-training phase of deep activity recognition models. Specifically, a generative pre-training of a deep model guides the discriminative training to better generalization solutions (Erhan et al. 2010). Clearly, the generative pre-training is almost ineffective for 1-layer networks. However, using the generative pre-training becomes more essential for the recognition accuracy of deeper activity recogni-

Table 2: Comparison of accuracy improvements due to the pre-training stage. Each layer consists of 1000 neurons.

EXPRIEMENT	# OF LAYERS	ACCURACY (%)
Generative & discriminative training	1	96.87
	3	97.75
	5	97.85
Discriminative training only	1	96.87
	3	96.46
	5	96.51

tion models, e.g., 5 layers.

Conclusions and Future Work

We investigated the problem of activity recognition using triaxial accelerometers. The proposed approach is superior to traditional methods of shallow networks with handcrafted features by using deep activity recognition models. The deep activity recognition models produce significant improvement to the recognition accuracy by extracting hierarchical features from triaxial acceleration data.

Even though this paper focuses on using triaxial accelerometers, the deep activity recognition models are general and can be applied with multiple types of sensors, e.g., gyroscopes and accelerometers simultaneously. Specifically, using multiple types of sensors results in multimodal data which is useful in accuracy-sensitive applications with complex activity set.

References

Altun, K., and Barshan, B. 2010. Human activity recognition using inertial/magnetic sensor units. In *Human Behavior Understanding*. Springer. 38–51.

Bächlin, M.; Plotnik, M.; Roggen, D.; Maidan, I.; Hausdorff, J. M.; Giladi, N.; and Tröster, G. 2010. Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine* 14(2):436–446.

Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H.; et al. 2007. Greedy layer-wise training of deep networks. *Advances in neural information processing systems* 19:153.

Berchtold, M.; Budde, M.; Gordon, D.; Schmidtke, H. R.; and Beigl, M. 2010. ActiServ: Activity recognition service for mobile phones. In *Proceedings of the International Symposium on Wearable Computers*, 1–8. IEEE.

Catal, C.; Tufekci, S.; Pirmit, E.; and Kocabag, G. 2015. On the use of ensemble of classifiers for accelerometer-based activity recognition. *Applied Soft Computing*.

Chen, L.; Hoey, J.; Nugent, C. D.; Cook, D. J.; and Yu, Z. 2012. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 42(6):790–808.

Dahl, G. E.; Yu, D.; Deng, L.; and Acero, A. 2012. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 20(1):30–42.

Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.-A.; Vincent, P.; and Bengio, S. 2010. Why does unsupervised pre-training help deep learning? *The Journal of Machine Learning Research* 11:625–660.

Fender, A.; MacPherson, W. N.; Maier, R.; Barton, J. S.; George, D. S.; Howden, R. I.; Smith, G. W.; Jones, B.; McCulloch, S.; Chen, X.; et al. 2008. Two-axis temperature-insensitive accelerometer based on multicore fiber Bragg gratings. *IEEE sensors journal* 7(8):1292–1298.

Hammerla, N. Y.; Kirkham, R.; Andras, P.; and Ploetz, T. 2013. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In *Proceedings of the International Symposium on Wearable Computers*, 65–68. ACM.

Hinton, G. E.; Osindero, S.; and Teh, Y.-W. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18(7):1527–1554.

Hinton, G. E. 2002. Training products of experts by minimizing contrastive divergence. *Neural computation* 14(8):1771–1800.

Hinton, G. E. 2012. A practical guide to training restricted Boltzmann machines. In *Neural Networks: Tricks of the Trade*. Springer. 599–619.

Khan, A. M.; Lee, Y.-K.; Lee, S. Y.; and Kim, T.-S. 2010. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine* 14(5):1166–1172.

Kwapisz, J. R.; Weiss, G. M.; and Moore, S. A. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12(2):74–82.

Lara, O. D., and Labrador, M. A. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* 15(3):1192–1209.

Mohamed, A.-R.; Dahl, G. E.; and Hinton, G. 2012. Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing* 20(1):14–22.

Parkka, J.; Ermes, M.; Korpipaa, P.; Mantyjarvi, J.; Peltola, J.; and Korhonen, I. 2006. Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine* 10(1):119–128.

Plötz, T.; Hammerla, N. Y.; and Olivier, P. 2011. Feature learning for activity recognition in ubiquitous computing. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, 1729.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature* 323(6088):533–536.

Salakhutdinov, R. 2015. Learning deep generative models. *Annual Review of Statistics and Its Application* 2(1):361–385.

Xu, W.; Zhang, M.; Sawchuk, A. A.; and Sarrafzadeh, M. 2012. Robust human activity and sensor location corecognition via sparse signal representation. *IEEE Transactions on Biomedical Engineering* 59(11):3169–3176.

Zappi, P.; Lombriser, C.; Stiefmeier, T.; Farella, E.; Roggen, D.; Benini, L.; and Tröster, G. 2008. Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. In *Wireless sensor networks*. Springer. 17–33.

Zheng, F.; Zhang, G.; and Song, Z. 2001. Comparison of different implementations of MFCC. *Journal of Computer Science and Technology* 16(6):582–589.